Int J Legal Med (2004) 118 : 83–89
DOI 10.1007/s00414-003-0421-4

ORIGINAL ARTICLE

S. Rand · M. Schürenkamp · C. Hohoff · B. Brinkmann

The GEDNAP blind trial concept part II. Trends and developments

Received: 28 October 2003 / Accepted: 25 November 2003 / Published online: 31 January 2004

Abstract This article presents a review of the developments in the GEDNAP blind trials over the period covering the past 10 years (1993–2003), demonstrating the changing approach to DNA investigations in the European community as a whole. The results of the trials also identify the most common types of error encountered which can also occur during routine DNA typing and ways of recognising such errors are suggested.

Keywords GEDNAP · Developments · Quality assurance · Quality control · Error recognition

Introduction

The introduction of a concise structural organisation and concept for the GEDNAP blind trials (Rand et al. 2002) has found broad positive resonance in the forensic community in most European countries. This seems to be due to a combination of several factors, such as increasing demands on laboratories as a result of database methods of criminal intelligence, the clear move towards accreditation and international collaboration and therefore harmonisation on a European-wide basis (Morling 2003; Morling et al. 2003). The continuing progress and development of new systems for forensic genetic identification necessitates a parallel progress in quality control and management at the laboratory, national and international levels which is mirrored in the progression of the structure and performance of the GEDNAP blind trials.

The results obtained in the initial phases of the GEDNAP trials involving single locus systems (SLS) and AmpFlps have already been reported in several publications (Bär et al. 1992; Puers et al. 1992; Brinkmann et al. 1993; Wiegand et al. 1995) but there have been no further publications of the results since the introduction of the most recent developments initially using STRs (Wiegand et al. 1993), then multiplex techniques (see e.g. Sparkes et al. 1996a, 1996b; Cotton et al. 2000), Y-chromosome STRs (see Gill et al. 2001; Gusmao et al. 2002) as well as the systems currently used for the German DNA database (Junge and Madea 2002; Junge et al. 2003; Klein et al. 2003).

This current publication cannot give detailed results obtained by the participating laboratories over the last 10 years due to the vast amount of information accumulated and also in order to maintain anonymity, but instead attempts to give a review of the most important developments and trends with regard to the quality of results and the logistics over this time period. It also attempts to analyse the types of errors which have been encountered during this period to pinpoint what we can learn from these mistakes.

The last results published were those of GEDNAP 4 and 5 (Wiegand et al. 1995) from the year 1992 whereby 2 trials were sent out each year to participants in accordance with the general recommendations of the German Medical Council (see e.g. Bundesärztekammer 1992). All subsequent trials as well as the current trials GEDNAP 26 and 27 now being held in 2003 continue to follow this established tradition and although this is not an international regulation applicable for all laboratories, it has been decided by consensus opinion of the participants to be an acceptable condition (see Rand et al. 2002).

Materials and methods

The material reviewed in this study is taken from the results obtained by laboratories participating in the GEDNAP blind trials between the years 1993–2003, encompassing GEDNAP trials 6–27. Any reference to a particular laboratory or country or other means which could lead to recognition have been omitted to strictly adhere to a policy of anonymity.

The results of the two trials for each year have been regularly presented on the occasion of the Stain Workshop (for details see http://www.gednap.de.vu) in the February of the year following the trials. All results have always been presented in an anonymous form in accordance with the published guidelines (Rand et al. 2002).

S. Rand (✉) · M. Schürenkamp · C. Hohoff · B. Brinkmann
Institut für Rechtsmedizin,
Universitätsklinikum Münster,
Röntgenstrasse 23, 48149 Münster, Germany
Tel.: +49-251-8355171, Fax: +49-251-8355158,
e-mail: rand@uni-muenster.de

Although the results of each individual set of trials are conveyed to the participants, it would seem to be useful to summarise the important points from all trials together in order to review the trends and developments which have occurred during this time period.

Results

There are a number of criteria which would seem to be of importance in charting the various stages and the important milestones in the development of the GEDNAP blind trials which are detailed in the following.

Developments in number of participants

The number of participants has shown a dramatic change during this period with an overall increase from 29 laboratories in 1993 (GEDNAP 6 and 7), with only 9 laboratories testing the system TH01, the first STR to be introduced, to the current status in 2003 of 160 laboratories for GEDNAP 26 and 27 (Table 1). There has always been a gradual increase over the whole period but in the last 3 years in parallel with the innovative introduction of the DNA database in Germany (and also in other European countries) and the decision to make GEDNAP the general standard to qualify as an input laboratory, there has been a sharp increase in the number of laboratories notably in the private sector. In addition, the decision of the ENFSI laboratories (Gill et al. 2000) to also use this facility for harmonisation of European forensic laboratories, has meant that the number of participants has shown a remarkable increase of nearly 100% over the last 3 years (Table 1).

The number of European countries involved has also shown a corresponding increase from 4 in 1993 to 31 in 2003 and reflects the growing concern for standardisation throughout Europe. This means that practically all countries in Europe are now represented by one or more laboratories.

Developments and landmarks in systems tested

In 1993 the first STR TH01 (TC11 Edwards et al. 1992) was introduced into the trials which were performed by 6 (GEDNAP 6) or 9 (GEDNAP 7) laboratories mostly using gel electrophoresis and silver staining. Over the subsequent trials other STRs were introduced, notably FXIIIB, FES and CD4, but were subsequently discontinued in 2000 (GEDNAP 20 and 21) due to the changing technologies and the development of new STR systems. Only TH01 and VWA have continued throughout and are still being used as part of multiplex kits.

In 1996 (GEDNAP 12 and 13) the first Y-chromosome STR DYS19 was introduced and in 1998 (GEDNAP 16 and 17) notably the STRs ACTBP2 and D21S11 were included although just as a pilot project involving selected laboratories. The main problem being the nomenclature and separation of the many structural and sequence variants (Schneider et al. 1998; Junge and Madea 2002). The ACTBP2 system is still considered to be too variable to be suitable for use in many national databases but does play a key role in the German database. With the advent of multiplex testing in 1999 (GEDNAP 18 and 19) all of the initial systems with the exception of TH01 and VWA had

Table 1 Details of the participation of laboratories and results in the trials GEDNAP 6/7 (1993) through GEDNAP 26/27 (2003)

GEDNAP (year)	Number of laboratories	Number of countries	Number of STRs tested	Size of stains tested (µl)	Number of tests	Error rate (%)
6 and 7 (1993)	29	4	1	5–300	104	0
8 and 9 (1994)	38	7	3	200	811	1.2
10 and 11 (1995)	48	9	5	20–200	2,550	2.1
12 and 13 (1996)	59	10	8	10–100	4,448	1.2
14 and 15 (1997)	72	11	8	5–50	5,184	1.6
16 and 17 (1998)	75	12	10	20–50	7,468	0.7
18 and 19 (1999)	78	12	21	5–50	11,409	0.5
20 and 21 (2000)	85	12	15	25	13,868	0.7
22 and 23 (2001)	122	28	17	10–25	21,743	0.5
24 and 25 (2002)	136	30	17	5–25	30,479	0.4
26 and 27 (2003)	160	31	30	5–25	n.a	n.a.

n.a. not available

Table 2 Current STR systems included in GEDNAP 26 and 27 (2003)

Autosomal STRs			Y-STRs	
TH01	D18S51	D13S317	DYS19	DYS437
VWA	D16S539	D7S820	DYS385	DYS438
FGA (Fibra)	D2S1338	Penta D	DYS389I/II	DYS439
D21S11	D19S433	Penta E	DYS390	YCAII
ACTBP2 (SE33)	TPOX	Amelogenin	DYS391	DXYS156-Y
D3S1358	CSF1PO		DYS392	
D8S1179	D5S818		DYS393	

been discontinued as they were no longer considered to be in the main frame of forensic analyses. All SLPs had long been discontinued and now the HLA-related marker kits and the last surviving Ampflp D1S80 were also dropped.

The number of systems now began to stabilise but still showed a steady increase due to the increasing number of multiplex kits available and also the number of markers included in each kit. Currently 18 autosomal STRs are included and just in the trials 24 and 25 (2002) and again in 26 and 27 (2003) there has been by popular demand a new influx of Y-STRs (Table 2).

Developments in success rates

The method of evaluating whether a laboratory has made an error has remained constant since the introduction of multiplex STR systems. This is independent of the method of detection or analysis and is based on four categories (Rand et al. 2002) of which only one (category 4) is classified as an error.

Despite the progress and developments made in forensic DNA typing over this time period, errors are still being made in the blind trials. Having said this the fact is that in nearly all of the trials carried out, most of the errors were made each time by a very few number of laboratories and of course compound errors such as the interchange of two samples caused a disproportional number of errors relative to the one mistake made when sampling the wrong test stain. However, this is not taken into consideration when calculating the error rate.

The zero error rate in GEDNAP 6/7 was due to the fact that very few laboratories participated, all were relatively experienced at that time and all used the same system of separation and detection, thus minimising any possible variation and therefore sources of error.

The error rates for the period 1994 through 1997 (GEDNAP 8/9–14/15) were in the region of 1–2% mainly due to problems of separation on polyacrylamide gels combined with structural variations within the STR systems (data not shown). The introduction of increased standardisation in 1998 (GEDNAP 16/17) linked to the availability of commercially available multiplex kits and ladders in 1999 (GEDNAP 18/19) has led to a sharp reduction of errors and the error rate was reduced to a relatively constant 0.4–0.7%.

However, the main outcome is that despite these continuous fluctuations and influx to the blind trials over this period, the error rate has remained remarkably constant.

Developments in types of error

The types of errors and the proportion of these errors has remarkably shown no great variation over the period in question despite the development in techniques and in particular the change to standardised multiplex technologies. However, considering the fact that the most common type of error has always been transcriptional errors followed by incorrect interpretation due to failing to recognise an error, these types of human error are to some extent obviously unavoidable under any of the prevailing circumstances.

The types of error found in the blind trials clearly show that human carelessness is the predominant source of error regardless of the technology used.

Two important problem areas have been identified over the 10 years of trials. Obviously one of the main problems is the transposition of samples in the laboratory which has occurred regularly in all trials carried out so far, albeit never by the same laboratory. But the fact that this is a recurrent problem would seem to indicate that generally more care should be taken to maintain the integrity of samples received in a laboratory.

A second problem has also regularly occurred with regard to transcription errors, i.e. writing a correct result incorrectly in the summary sheet supplied. This problem will be dealt with in more detail in the discussion.

Incorrect results have in the past also to some extent been associated with particular types of stains, such as cigarette butts and mixtures of body fluids as well as hairs. The latter have since been discontinued because of the inconsistency of the amount of DNA present and the resulting substantial problem of reproducibility within and between laboratories.

There are of course, certain situations where it is not possible to differentiate between a minor allele (signal) in a mixture and, for example a stutter band from an adjacent major signal. This is however not usually a major cause for concern if this possibility can be recognised as such. It would simple be recorded as inconclusive and would not be (falsely) reported.

Having said this there are two problems which occurred during the most recent trials which were associated to some extent with the reliance on automatic allele-calling and a lack of manual control of the computer print-outs.

The first example concerns long alleles in the Fibra system (FGA). Stain 2 in GEDNAP 24 was a 50:50 mixture of blood samples, one of which had the FGA allele 42.2. This allele is included in the allelic ladders of some commercially available kits (e.g. Powerplex from Promega Madison WI, SGM from ABI Foster City CA, MPX3 from Serac, Bad Homburg, Germany) but not in others (e.g. AmpFlSTR Profiler ABI, MPX2, Serac Bad Homburg) and in some instances not only appears as an off-ladder allele but also overlaps into the neighbouring system (Fig. 1). This caused some confusion among participants but in the majority of cases was recognised and correctly assigned. However, in some instances the long 42.2 allele was present but not recognised and in some other instances was recognised but incorrectly assigned to another system.

The second example concerned stain 5 in GEDNAP 25 which again was a 50:50 mixture of blood samples containing the TH01 alleles 8.3, 9, 9.3 and 10. Again this was recognised in the vast majority of cases as being a mixture with two allele pairs each showing 1 bp differences. In many cases (see Fig. 2) the automatic system failed to

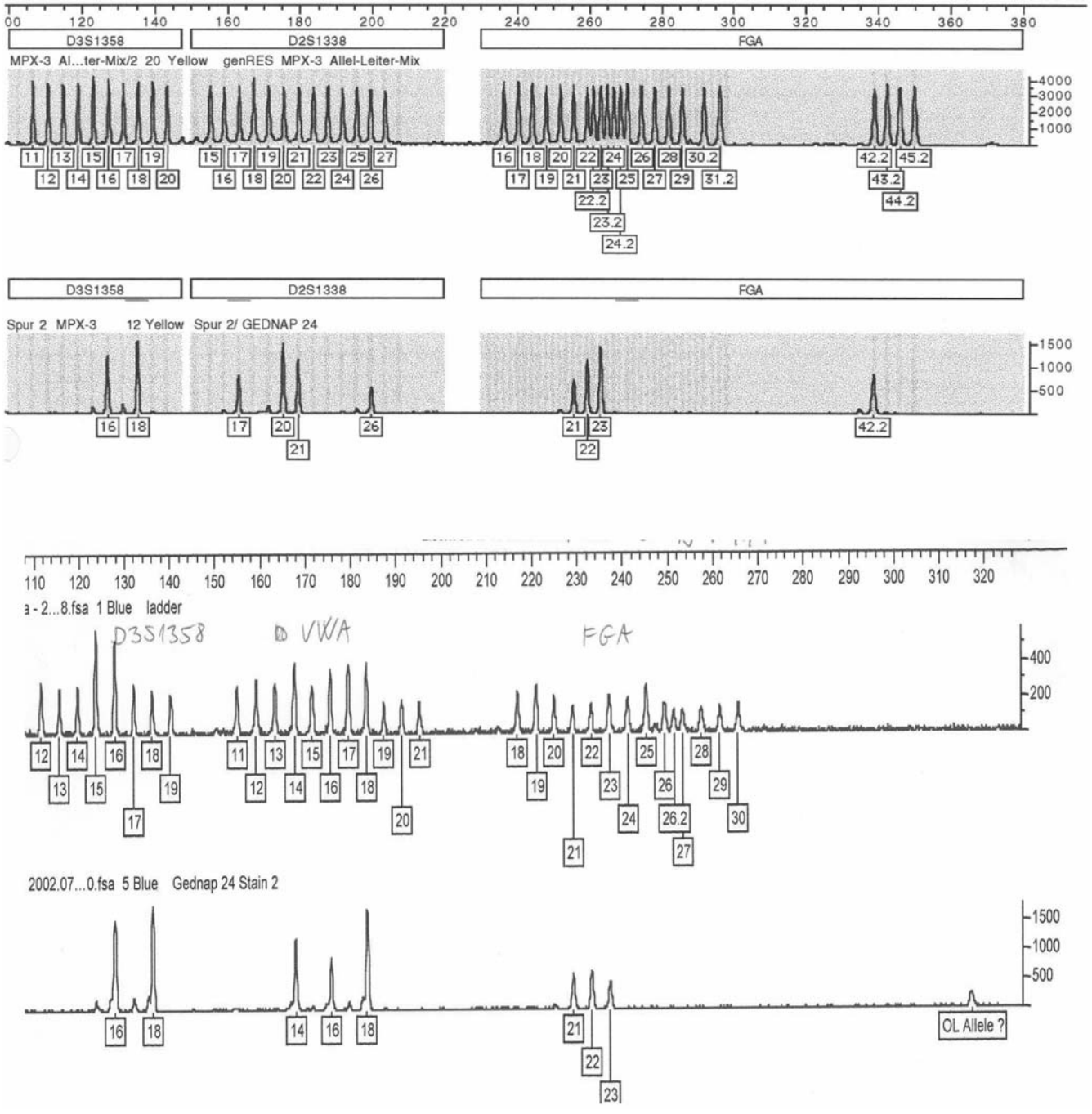

Fig. 1 *Upper panels* Correct typing of FGA 42.2 allele using the MPX3 multiplex kit (Serac, Bad Homburg) and *lower panels* as an off-ladder allele using Profiler (ABI)

recognise all the alleles and a manual check and correction became necessary.

Developments in investigated material

There have been no major changes in the approach of the GEDNAP trials over the period in question. The stains sent to participants have always reflected the current state-of-the-art and problem areas at the particular time of the trial, always with respect to the consensus of the participants and according to the decisions of the Stain Commission (see http://www.spurenkommission.de.vu). Several aspects have arisen over the course of the trials which have affected the type of material sent out.

The inclusion of mixtures of body fluids in these trials has always been a major aspect and is reflected in the fact that mixtures have been an integral component since the very beginning. However, changes in the sensitivity of detection, and development of male-specific Y-chromosome markers in particular has led to an improvement in the capability of detection of minor male components in a mixture.

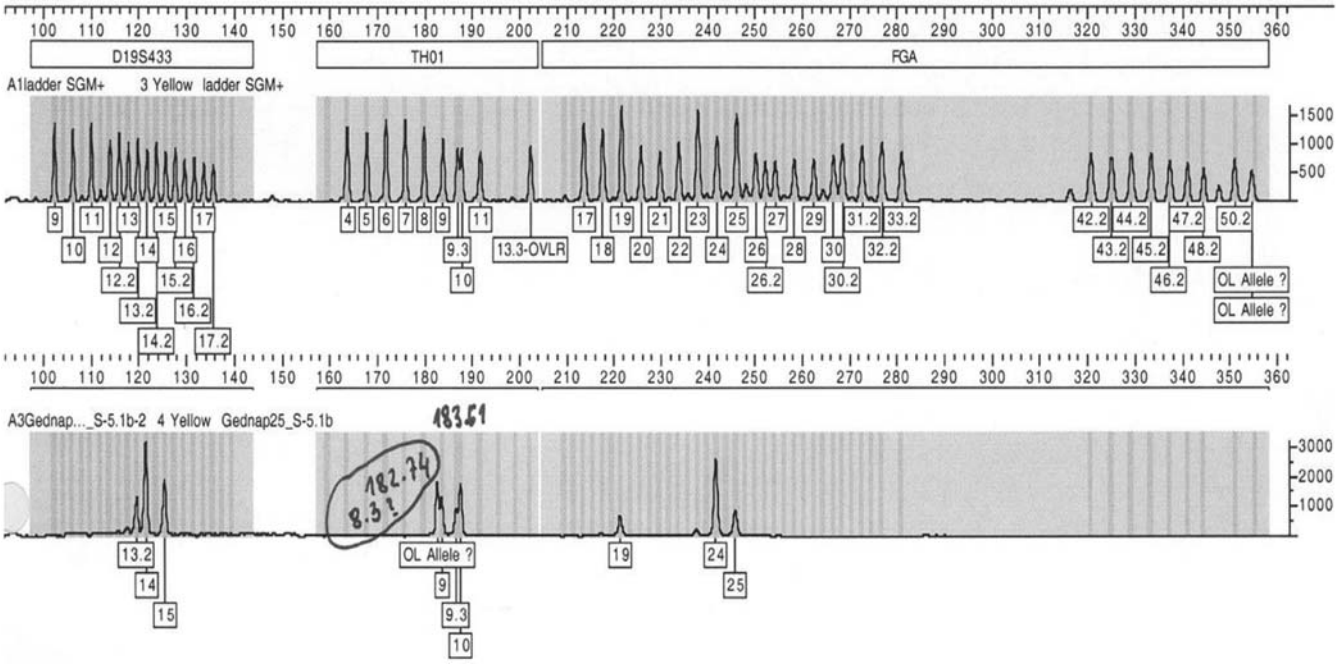

Fig. 2 Manual correction of the TH01 results obtained for stain 5 using the AmpFlSTR Profiler system (ABI)

During the course of the blind trials various stains have been sent most of which have caused no problems (but see also previous section) to laboratories routinely dealing with forensic casework.

Developments in logistics

The dramatic increase in the number of participants has necessitated a streamlining of the logistics involved in the production and transport of samples. However, through experience gained during this period, the development an optimalised procedure for dealing with incoming results and standardisation of the way in which results are submitted, the workload has remained relatively constant. This also minimises the introduction of errors during registration and evaluation of results submitted. A standard form has been introduced for the recording and submission of results which are then logged into a standard computer programme. Each step of recording and entry of results is checked by two independent members of staff in order to immediately recognise incorrect entries.

Discussion

One of the main aspects of the trials has always been and will continue to be that the samples sent out correspond as near as possible to real casework situations. This includes such stains as cigarette butts stained with saliva, artificial mixtures of various body fluids which could normally be encountered and straightforward mixtures of 2 blood samples which are also very common in forensic casework. There are no tricks involved and even though rare alleles are sometimes included this represents the reality of everyday casework. Laboratories must be prepared for the fact that depending on throughput, rare alleles will be encountered in all STR systems with more or less regularity during routine testing and a system must be implemented in each laboratory for the recognition and detection of such rare alleles or other abnormalities.

It cannot be acceptable that a modern day forensic science laboratory operates on the principle that rare alleles occur so infrequently that they do not therefore need to be identified as such.

With the modern technologies now available, every laboratory should be capable of detecting a 1 bp difference when assigning alleles. Even if results are not immediately conclusive, there must be safeguards so that for example an individual with a 9.3 or 10 allele in the TH01 system can be recognised. This is especially important for the system ACTBP2 (SE33) for laboratories contributing to the German database because 0.3 and 0.1 variants are known to occur and although not common with frequencies less than 1%, can be a regular finding in high volume casework. It is therefore absolutely necessary that these variants can be identified to avoid false DNA profile information being included in the database.

Although simulated case situations have been considered, the practicality of each participant receiving identical and reproducible sample quality has always been a limiting factor for this approach. The next best and not necessarily less effective approach of carefully prepared and measured samples and mixtures has proven to be a valid and reproducible method.

This is reflected in the results obtained by the vast majority of participating laboratories who consistently and reproducibly produce correct results. A variety of mixures

of various body fluids and combinations of alleles have been included in the trials over the period in question. Important is that the test results should reflect what is achievable according to the current state-of-the-art.

Obviously some laboratories will have more experience and available up-to-date technology than others. This will never change due to the different economic and political status of legal systems throughout Europe in general. But wrong results should never be obtained regardless of these limiting factors and forensic science is duty bound to produce results of the best possible standard to maintain its obligation to justice and society. It is the moral obligation and the fulfillment of a European quality control system to maintain standards which support a true democratic and impartial legal system.

The sharp increase in the number of participating laboratories has meant that with each new trial, at least over the past three trials, there has always been a number of new participating laboratories who are not as familiar with the GEDNAP procedure, for example in terms of interpretation and presentation of results, as other laboratories who have participated for many years. Also as an example, the idiosyncrasies of particular legal systems may demand that all DNA allelic signals detected, however small and regardless of the reason, as for example stutter bands, must be included in the presentation of results in all cases. This is not taken into consideration when evaluating the results but should be borne in mind by the participants when the results are made known.

To be fair to the participating laboratories and perhaps to put the error rate into a more realistic context, the system of submitting the results of the trials is not completely satisfactory and could per se be the cause of errors which under the normal working routine would not have occurred.

The method of submission of results is that a standard form is sent to each participant which must be filled out by hand so as to standardise the way in which they are presented for evaluation. As in many instances a laboratory has developed its own automated custom-made approach to checking the interpretation and report writing, in these cases this would not be the way it would normally be performed and brings in a further possible source of error. Unfortunately, due to this interlaboratory heterogeneity of presentation of results and to the sheer volume of results which need to be evaluated, there is no way of being able to satisfy the demands of all the different methods simultaneously while still maintaining a robust and practical way of evaluating the results.

There are a number of other systems and technologies which are at present being used in special casework situations and are not yet included in GEDNAP such as mitochondrial DNA (mtDNA) (e.g. Baer et al. 2000; Tully et al. 2001) or others which are being validated by laboratories involved in research, such as SNPs (Gill 2001; Brandstätter et al. 2003, and for further up-to-date information on the forensic application of SNPs see Doutremépuich and Morling 2004) or X chromosomal STRs (Szibor et al. 2003; Wiegand et al. 2003). This has always been the case and is why forensic science has progressed so rapidly over the last two decades or so. But there are limitations to all systems of quality control and in order to maintain a high standard it is necessary to limit the systems under consideration to those which currently form the routine basis of forensic DNA testing in the courtroom setting.

New developments in the forensic application of DNA technologies will always be considered for inclusion in the GEDNAP blind trials providing they represent a majority opinion.

In conclusion, the changing face of DNA technology and the associated progress in forensic investigations is reflected in the increased standardisation and reproducibility and most important with the very low error rate in the GEDNAP blind trials which has been achieved by the vast majority of participating laboratories.

During the development of the trials over the last decade since the introduction of STRs many lessons have been learnt which have led to improvements not only in the trials themselves, but also have achieved the goal of improved quality of results within the forensic community as a whole.

References

Bär W, Brinkmann B, Germann U et al. (1992) GEDNAP (German-DNA-Profiling-Group) II. Projekt zur Standardisierung und Methodenharmonisierung in der forensischen DNA-Technologie. Rechtsmedizin 3:6–9

Baer W, Brinkmann B, Budowle B et al. (2000) DNA Commission of the International Society for Forensic Genetics guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196

Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Legal Med (in press) DOI: s00414-003-0395-2

Brinkmann B, Bär W, Feuerbach M et al. (1993) GEDNAP III – Dritter Spurenringversuch der Deutschen Gesellschaft für Rechtsmedizin. Rechtsmedizin 3:33–39

Bundesärztekammer (1992) Richtlinien der Bundesärztekammer zur Qualitätssicherung in der Immunhämatologie. Dtsch Arztebl 89:479–482

Cotton EA, Allsop RF, Guest JL et al. (2000) Validation of the AMPF/STR SGM plus system for use in forensic casework. Forensic Sci Int 112:151–161

Doutremépuich C, Morling N (eds) (2004) Progress in Forensic Genetics 10. Elsevier, Amsterdam

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241–253

Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 114:204–210

Gill P, Sparkes R, Fereday L, Werrett DJ (2000) Report of the European Network of Forensic Science Institutes (ENFSI): formulation and testing of principles to evaluate STR multiplexes. Forensic Sci Int 108:1–29

Gill P, Brenner C, Brinkmann B et al. (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y chromosome STRs. Int J Legal Med 114:305–309

Gusmao L, Alves C, Beleza S et al. (2002) Forensic evaluation and population data on the new Y-STRs DYS434, DYS437, DYS438, DYS439 and GATA A10. Int J Legal Med 116:139–147

Junge A, Madea B (2002) Forensische Spurenkunde. Genomxpress 1:4–5

Junge A, Lederer T, Braunschweiger G et al. (2003) Validation of the multiplex kit *genRES*MPX-2 for forensic casework analysis. Int J Legal Med (in press) DOI: s00414-003-0385-4

Klein R, Braunschweiger G, Junge A, Wiegand P (2003) A very long ACTBP2 (SE33) allele. Int J Legal Med 117:235–236

Morling N (2003) International recommendations for paternity testing standards. Int J Legal Med 117:1

Morling N, Allen R, Carrecedo A et al. (2003) Paternity testing commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. Int J Legal Med 117:51–61

Puers C, Wiegand P, Bär W et al. (1992) GE-DNA-P (German DNA Profiling Group)-Studie. Rechtsmedizin 2:99–103

Rand S, Schürenkamp M, Brinkmann B (2002) The GEDNAP blind trial concept. Int J Legal Med 117:199–206

Schneider HR, Rand S, Schmitter H, Weichhold G (1998) ACTBP2-nomenclature recommendations of GEDNAP. Int J Legal Med 111:97–100

Sparkes R, Kimpton C, Watson S et al. (1996a) The validation of a 7-locus multiplex STR test for use in forensic casework I. Mixtures, ageing, degradation and species studies. Int J Legal Med 109:186–194

Sparkes R, Kimpton C, Gillbard S et al. (1996b) The validation of a 7-locus multiplex STR test for use in forensic casework II. Casework studies and success rates. Int J Legal Med 109:195–204

Szibor R, Krawczak M, Hering S et al. (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74

Tully G, Baer W, Brinkmann B et al. (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83–91

Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE33 and TC11. Int J Legal Med 105:315–320

Wiegand P, Ambach E, Augustin C et al. (1995) GEDNAP IV and V. The 4th and 5th stain blind trials using DNA technology. Int J Legal Med 108:79–84

Wiegand P, Berger B, Edelmann J et al. (2003) Population genetic comparisons of three X-chromosomal STRs. Int J Legal Med 117:62–65